1.0 ACCURACY FOR THE PREDICTION OF GHS ACUTE ORAL TOXICITY CATEGORY

This analysis of accuracy for the prediction of GHS acute oral toxicity categories (UN 2005) predicted by the 3T3 and NHK NRU test methods was performed using all the available IC_{50} data: 70 substances for the 3T3 NRU and 71 substances for the NHK NRU. Of the 72 substances tested in the study, carbon tetrachloride and methanol were excluded from the 3T3 NRU analysis and methanol was excluded from the NHK NRU analysis because no laboratory attained sufficient toxicity in any test for the calculation of an IC_{50}.

1.1 Prediction of Toxicity Category by the 3T3 and NHK NRU Test Methods Using the RC Millimole Regression

Table 1-1 shows the concordance of the observed and predicted GHS acute oral toxicity categories (UN 2005) for each *in vitro* cytotoxicity test method using the geometric mean IC_{50} values (of the three laboratories) in the RC millimole regression, log LD_{50} (mmol/kg) = 0.435 x log IC_{50} (mM) + 0.625. Accuracy is the agreement of the category predictions with those based on the initial rodent LD_{50} values used for selected substances for testing (in Table 3-2 of the BRD). Substances for which the *in vitro* toxicity category prediction does not match the *in vivo* determined toxicity category are considered discordant substances for the GHS toxicity category predictions.

For the 3T3 NRU test method, the toxicity category was underpredicted for 29 (56%) and overpredicted for 22 (43%) of the 51 discordant substances. For the NHK NRU test method, toxicity was underpredicted for 28 (54%) and overpredicted for 24 (46%) of the 52 discordant substances. The fact that there were more substances that were underpredicted for toxicity is consistent with the RC substances chosen for testing. Figure 3-1 of the BRD shows that most of the selected RC substances are below the RC millimole regression line. Thus, the RC is expected to predict a higher LD_{50} (i.e., lower toxicity).

Table 1-1 Prediction of GHS Toxicity Category[1] by the 3T3 and NHK NRU Test Methods and the RC Millimole Regression

Initial Rodent LD_{50}[2]	3T3 NRU-Predicted Toxicity Category						Total	Accuracy	Toxicity Overpredicted	Toxicity Underpredicted
	< 5	5 – 50	50 – 300	300 – 2000	2000 – 5000	> 5000				
< 5	0	3	1	8	0	0	12	0%	0%	100%
5 –50	0	2	5	4	1	0	12	17%	0%	83%
50 – 300	0	0	5	7	0	0	12	42%	0%	58%
300 – 2000	0	0	1	11	0	0	12	92%	8%	0%
2000 – 5000	0	0	0	11	0	0	11[3]	0%	100%	0%
> 5000	0	0	0	7	3	1	11[4]	9%	91%	0%
Total	0	5	12	48	4	1	70	27%	31%	41%
Predictivity	0%	40%	42%	23%	0%	100%				
Category Underpredicted	0%	0%	8%	38%	75%	0%				
Category Overpredicted	0%	60%	50%	40%	25%	0%				
Initial Rodent LD_{50}	**NHK NRU-Predicted Toxicity Category**						**Total**	**Accuracy**	**Toxicity Overpredicted**	**Toxicity Underpredicted**
	< 5	**5 – 50**	**50 – 300**	**300 – 2000**	**2000 – 5000**	**> 5000**				
< 5	0	1	3	7	1	0	12	0%	0%	100%
5 – 50	0	4	7	1	0	0	12	33%	0%	67%
50 – 300	0	1	4	7	0	0	12	33%	8%	58%
300 – 2000	0	0	1	10	1	0	12	83%	8%	8%
2000 – 5000	0	0	0	10	1	0	11[3]	9%	91%	0%
> 5000	0	0	1	6	5	0	12	0%	100%	0%
Total	0	6	16	41	8	0	71	27%	34%	39%
Predictivity	0%	67%	25%	24%	13%	0%				
Category Underpredicted	0%	17%	13%	39%	63%	0%				
Category Overpredicted	0%	17%	63%	37%	25%	0%				

[1]GHS-Globally Harmonized System of Classification and Labelling of Chemicals with LD_{50} in mg/kg (UN 2005). The RC millimole regression is log LD_{50} (mmol/kg) = log IC_{50} (mM) X 0.435 + 0.625. Numbers in table represent number of substances.

[2]Initial rodent LD_{50} values from Table 3-2 of the BRD.

[3]Carbon tetrachloride excluded because no laboratory attained sufficient toxicity for the calculation of an IC_{50}.

[4]Methanol excluded because no laboratory attained sufficient toxicity for the calculation of an IC_{50}.

1.2 Prediction of Toxicity Category by the 3T3 and NHK NRU Test Methods Using the RC Rat-Only Weight Regression

Table 1-2 shows the concordance of the observed and predicted GHS acute oral toxicity categories for each test method using the geometric mean IC_{50} values (of the three laboratories) and the RC rat-only weight regression from Table 6-2 of the BRD. The regression formula for the RC rat-only weight regression was log LD_{50} (mg/kg) = log IC_{50} (μg/mL) x 0.372 + 2.024. Accuracy is the agreement of the *in vitro* NRU cytotoxicity GHS toxicity category predictions with those based on the reference rat oral LD_{50} values from Table 4-2 of the BRD.

The two *in vitro* NRU cytotoxicity test methods over- and under-predicted the GHS toxicity category for a similar number of substances, compared with the GHS toxicity categories for the reference LD_{50} values in Table 4-2 of the BRD. For the 3T3 NRU test method, the GHS toxicity category of 23 (48%) of 48 discordant substances was overpredicted and the GHS toxicity category of 25 (52%) substances was underpredicted. For the NHK NRU test method, the GHS toxicity category of 26 (53%) of 49 discordant substances was overpredicted and the toxicity of 23 (47%) discordant substances was underpredicted.

Table 1-2 Prediction of GHS Toxicity Category[1] by the 3T3 and NHK NRU Test Methods and the RC Rat-Only Weight Regression

Reference Rodent LD_{50}[2]	3T3 NRU Predicted Toxicity Category						Total	Accuracy	Toxicity Overpredicted	Toxicity Underpredicted
	< 5	5 – 50	50 – 300	300-2000	2000-5000	> 5000				
< 5	0	0	2	5	0	0	7	0%	0%	100%
5 – 50	0	2	5	5	0	0	12	17%	0%	83%
50 – 300	0	0	4	8	0	0	12	33%	0%	67%
300 – 2000	0	1	3	12	0	0	16	75%	25%	0%
2000 – 5000	0	0	0	6	4	0	10[3]	40%	60%	0%
> 5000	0	0	0	6	7	0	13[4]	0%	100%	0%
Total	0	3	14	42	11	0	70	31%	33%	36%
Predictivity	0%	67%	29%	29%	36%	0%				
Category Underpredicted	0%	0%	50%	43%	0%	0%				
Category Overpredicted	0%	33%	21%	29%	64%	0%				
Reference Rodent LD_{50}[2]	**NHK NRU Predicted Toxicity Category**						**Total**	**Accuracy**	**Toxicity Overpredicted**	**Toxicity Underpredicted**
	< 5	**5 – 50**	**50 – 300**	**300 – 2000**	**2000 – 5000**	**> 5000**				
< 5	0	1	2	4	0	0	7	0%	0%	100%
5 – 50	0	2	5	5	0	0	12	17%	0%	83%
50 – 300	0	1	5	6	0	0	12	42%	8%	50%
300 – 2000	0	1	2	13	0	0	16	81%	19%	0%
2000 – 5000	0	0	0	9	1	0	10[3]	10%	90%	0%
> 5000	0	0	0	7	6	1	14	7%	93%	0%
Total	0	5	14	44	7	1	71	31%	37%	32%
Predictivity	0%	40%	36%	30%	14%	0%				
Category Underpredicted	0%	20%	50%	34%	0%	0%				
Category Overpredicted	0%	25%	50%	33%	0%	0%				

[1]Globally Harmonized System of Classification and Labelling of Chemicals with LD_{50} in mg/kg (UN 2005). The RC rat-only weight regression is log LD_{50} (mg/kg) = log IC_{50} (μg/mL) X 0.372 + 2.024.

77 [2]Reference rodent LD_{50} values from Table 4-2 of the BRD.
78 [5]Carbon tetrachloride excluded because no laboratory attained sufficient toxicity for the calculation of an IC_{50}.
79 [6]Methanol excluded because no laboratory attained sufficient toxicity for the calculation of an IC_{50},
80

1.3 Prediction of Toxicity Category by the 3T3 and NHK NRU Test Methods with the RC Rat-Only Weight Regression Excluding Substances with Specific Mechanisms of Toxicity

Table 1-3 shows the concordance of the observed and predicted GHS acute oral toxicity categories for each *in vitro* NRU test method using the geometric mean IC_{50} values (of the three laboratories) and the RC rat-only weight regression excluding substances with specific mechanisms of action (see **Table 6-2**). The formula for this regression was log LD_{50} (mg/kg) = log IC_{50} (μg/mL) x 0.357 + 2.194. Accuracy is the agreement of the *in vitro* predicted GHS toxicity categories with those based on the reference rat oral LD_{50} values from Table 4-2 of the BRD.

The NHK NRU test method had four more discordant substances than the corresponding assay using 3T3 cells when the IC_{50} results were applied to the RC rat-only weight regression excluding substances with specific mechanisms of toxicity. For the 3T3 NRU test method, the GHS toxicity category of 16 (38%) of 42 discordant substances was overpredicted while the toxicity of 26 (62%) of 42 discordant substances was underpredicted compared with the *in vivo* GHS toxicity categories for the reference LD_{50} values in **Table 4-2** of the BRD. For the NHK NRU test method, the toxicity of 21 (46%) of 46 discordant substances was overpredicted while the toxicity of 25 (53%) of 46 discordant substances was underpredicted.

Table 1-3 Prediction of GHS Toxicity Categories[1] by the 3T3 and NHK NRU with the RC Rat-Only Weight Regression Excluding Substances with Specific Mechanisms of Toxicity

Reference Rodent LD_{50}[2]	3T3 NRU Predicted Toxicity Category						Total	Accuracy	Toxicity Overpredicted	Toxicity Underpredicted
	< 5	5 – 50	50 – 300	300-2000	2000-5000	> 5000				
< 5	0	0	2	5	0	0	7	0%	0%	100%
5 – 50	0	2	4	6	0	0	12	17%	0%	83%
50 – 300	0	0	3	9	0	0	12	25%	0%	75%
300 – 2000	0	1	1	14	0	0	16	88%	13%	0%
2000 – 5000	0	0	0	4	6	0	10[3]	60%	40%	0%
> 5000	0	0	0	6	4	3	13[4]	23%	77%	0%
Total	0	3	10	44	10	3	70	40%	23%	37%
Predictivity	0%	67%	30%	32%	60%	0%				
Category Underpredicted	0%	0%	60%	45%	0%	0%				
Category Overpredicted	0%	33%	10%	23%	40%	0%				
Reference Rodent LD_{50}[2]	**NHK NRU Predicted Toxicity Category**						**Total**	**Accuracy**	**Toxicity Overpredicted**	**Toxicity Underpredicted**
	< 5	**5 – 50**	**50 – 300**	**300 – 2000**	**2000 – 5000**	**> 5000**				
< 5	0	0	2	5	0	0	7	0%	0%	100%
5 – 50	0	2	5	5	0	0	12	17%	0%	83%
50 – 300	0	1	4	7	0	0	12	33%	8%	58%
300 – 2000	0	1	1	13	1	0	16	81%	13%	6%
2000 – 5000	0	0	0	6	4	0	10[3]	40%	60%	0%
> 5000	0	0	0	5	7	2	14	14%	86%	0%
Total	0	4	12	41	12	2	71	35%	30%	35%
Predictivity	0%	50%	33%	32%	33%	100%				
Category Underpredicted	0%	0%	58%	41%	8%	0%				
Category Overpredicted	0%	50%	8%	27%	58%	0%				

[1]Globally Harmonized System of Classification and Labelling of Chemicals with LD_{50} in mg/kg (UN 2005). The RC rat-only weight regression excluding substances with specific mechanisms of toxicity is log LD_{50} (mg/kg) = log IC_{50} (μg/mL) X 0.357 + 2.194.

[2]Reference rodent LD_{50} values from Table 4-2 of the BRD.

[5]Carbon tetrachloride excluded because no laboratory attained sufficient toxicity for the calculation of an IC_{50}.

[6]Methanol excluded because no laboratory attained sufficient toxicity for the calculation of an IC_{50},

1.4 Summary of the Regressions Evaluated

Table 1-4 summarizes the regressions evaluated for accuracy in predicting the GHS acute oral toxicity categories (UN 2005), and the proportion of *in vitro* predicted discordant substances for each GHS toxicity category. Accuracy for both NRU cytotoxicity test methods was the same (27% for the RC, 31% for the RC rat-only) for the regressions evaluated except for the RC rat-only weight regression excluding substances with specific mechanisms of toxicity. For the latter regression, the accuracy of the 3T3 NRU test method was higher than that for the NHK NRU (40% vs. 35%, respectively). The proportion of discordant substances for the 3T3 NRU test method was also the same as that for the NHK NRU test method for the RC (73%) and RC rat-only (69%) regressions. The 3T3 NRU had a lower proportion of discordant substances for the RC rat-only weight regression excluding substances with specific mechanisms of toxicity (60% for the 3T3 NRU vs. 65% for the NHK NRU).

Table 1-4 Comparison of Regressions and *In Vitro* NRU Test Methods for Performance in Predicting GHS[a] Toxicity Categories

Regression	N[b]	Adjusted R^2	Accuracy	Discordant Substances[c]
RC –millimole units	347	0.450[d]	3T3 – 27% NHK – 27%	3T3- 51/70 (73%) NHK – 52/71 (73%)
RC rat only –weight units[e]	282	0.322	3T3 – 31% NHK – 31%	3T3- 48/70 (69%) NHK – 49/71 (69%)
RC rat only excluding substances with specific mechanisms of action – weight units[e]	232	0.353	3T3 – 40% NHK – 35%	3T3- 42/70 (60%) NHK – 46/71 (65%)

[a]Globally Harmonized System of Classification and Labelling of Chemicals with LD_{50} in mg/kg (UN 2005).
[b]Number of substances used in regression.
[c]Proportion of substances evaluated.
[d]Calculated from RC data (i.e., not reported by Halle [1998]).
[e]From Table 6-1 of the BRD.

The highest accuracy for both *in vitro* NRU cytotoxicity test methods was attained when using the RC rat only weight regression excluding substances with specific mechanisms of action. The accuracy for the 3T3 NRU test method was 40%, which was greater than the accuracy of the 3T3 NRU with the RC millimole regression (27%) and with the RC rat-only weight regression (31%). The accuracy for the NHK NRU test method was 35% for the RC

rat-only weight regression excluding substances with specific mechanisms of toxicity, 27% with the RC millimole regression, and 31% with the RC rat-only weight regression.

1.5 Alternate Accuracy Analysis for the RC Millimole Regression

This analysis of accuracy for the prediction of GHS acute oral toxicity categories (UN 2005) by the 3T3 and NHK NRU test methods was performed using the same IC_{50} data as used for the analyses above (70 substances for the 3T3 NRU and 71 substances for the NHK NRU). However, the *in vivo* GHS categories for this analysis are based on the reference LD_{50} values presented in Table 4-2 of the BRD rather than the initial LD_{50} values used to select the substances for testing (in Table 3-2). The analyses presented in **Table 1-1** used the initial LD_{50} values to determine the *in vivo* GHS acute oral toxicity categories.

For the 3T3 NRU test method, the toxicity category was underpredicted for 24 (49%) and overpredicted for 25 (51%) of the 49 discordant substances. For the NHK NRU test method, toxicity was underpredicted for 22 (44%) and overpredicted for 28 (56%) of the 50 discordant substances.

Table 1-5 Prediction of GHS Toxicity Category[1] by the 3T3 and NHK NRU Test Methods and the RC Millimole Regression Using Reference LD_{50} Values for *In Vivo* GHS Categories

Reference Rodent LD_{50}[2]	3T3 NRU-Predicted Toxicity Category						Total	Accuracy	Toxicity Overpredicted	Toxicity Underpredicted
	< 5	5 – 50	50 – 300	300 – 2000	2000 – 5000	> 5000				
< 5	0	2	0	5	0	0	7	0%	0%	100%
5 –50	0	2	5	4	1	0	12	17%	0%	83%
50 – 300	0	0	5	7	0	0	12	42%	0%	58%
300 – 2000	0	1	2	13	0	0	16	81%	19%	0%
2000 – 5000	0	0	0	10	0	0	10[3]	0%	100%	0%
> 5000	0	0	0	9	3	1	13[4]	8%	92%	0%
Total	0	5	12	48	4	1	70	30%	36%	34%
Predictivity	0%	40%	42%	27%	0%	100%				
Category Underpredicted	0%	40%	42%	33%	25%	0%				
Category Overpredicted	0%	20%	17%	40%	75%	0%				
Reference Rodent LD_{50}	**NHK NRU-Predicted Toxicity Category**						**Total**	**Accuracy**	**Toxicity Overpredicted**	**Toxicity Underpredicted**
	< 5	**5 – 50**	**50 – 300**	**300 – 2000**	**2000 – 5000**	**> 5000**				
< 5	0	1	2	4	0	0	7	0%	0%	100%
5 – 50	0	3	5	3	1	0	12	25%	0%	75%
50 – 300	0	1	6	5	0	0	12	50%	8%	42%
300 – 2000	0	1	2	12	1	0	16	75%	19%	6%
2000 – 5000	0	0	0	10	0	0	10[3]	0%	100%	0%
> 5000	0	0	1	7	6	0	14	0%	100%	0%
Total	0	6	16	41	8	0	71	30%	39%	31%
Predictivity	0%	50%	38%	29%	0%	0%				
Category Underpredicted	0%	17%	44%	29%	25%	0%				
Category Overpredicted	0%	17%	63%	37%	25%	0%				

[1]GHS-Globally Harmonized System of Classification and Labelling of Chemicals with LD_{50} in mg/kg (UN 2005). The RC millimole regression is log LD_{50} (mmol/kg) = log IC_{50} (mM) X 0.435 + 0.625. Numbers in table represent number of substances.

[2]Reference rodent LD_{50} values from Table 4-2 of the BRD.

[3]Carbon tetrachloride excluded because no laboratory attained sufficient toxicity for the calculation of an IC_{50}.

[4]Methanol excluded because no laboratory attained sufficient toxicity for the calculation of an IC_{50}.